

# 数据行为的概念界定与模型构建\*

■ 李文琦 张鹏翼

北京大学信息管理系 北京 100871

**摘要:** [目的/意义] 在数据密集型科研范式下,数据已经成为各类学术活动的基础,数据的管理和共享等实践也逐渐成为政策制定者、科研机构、数据服务提供者以及科研人员关注的焦点。目前,已有研究涉及到对科研相关的数据活动和实践、科研人员的态度和行为等层面的分析,然而尚未明确提出“数据行为”的概念。[方法/过程] 通过对相关文献的梳理整合,分析数据的性质及特征,阐述数据行为概念的必要性,并借鉴信息行为领域的相关理论提出个体视角下的“数据行为”概念,构建数据行为模型和概念框架。[结果/结论] 提出的个体视角下的数据行为模型概括数据行为的一般流程和活动,包括数据需求、数据收集行为、数据管理行为以及数据发表、共享、署名及引用行为;科研语境中的数据行为概念框架揭示对数据行为产生影响的各种因素,包括知识基础设施、科研情境和科研人员个体因素。该模型可以为科研人员数据行为的实证研究、数据政策制定、数据基础设施建设以及数据工具的设计提供理论基础和实践建议。

**关键词:** 数据行为 研究数据 数据实践 数据密集型研究 信息行为

**分类号:** G201

**DOI:** 10.13266/j.issn.0252-3116.2022.23.004

## 1 引言

随着大数据时代的到来,各学科、领域都面临着大规模数据的分析、管理和长期保存问题。数据基础设施的建设以及云计算、人工智能等技术的发展也推动科学研究转向数据密集型的第四范式<sup>[1]</sup>。一方面,新兴的数据科学蓬勃发展;另一方面,社会科学和人文研究也更多地通过数据驱动的方法来辅助资料分析,并发现新的研究问题<sup>[2]</sup>。数据已经成为各学科各类学术活动的基础。

英国研究委员会提出的 e-Science 计划<sup>[3]</sup>、美国国家科学基金会提出的网络基础设施项目 (Cyberinfrastructure Program)<sup>[4]</sup> 以及从自然科学延伸到社会科学及人文的各项议题<sup>[5-6]</sup> 促进了对数据密集型研究的基础设施建设以及数据的长期保存和管护的关注。已有研究从数据生命周期的角度对数据产生、处理分析、长期保存和重复利用过程中的最佳工作实践与问题进行了梳理<sup>[7-9]</sup>。随着开放科学的提出和发展<sup>[10]</sup>,科研管理和资助机构、学术期刊及出版商等都实施了一系

列举措来促进数据公开和数据共享<sup>[11]</sup>。例如,很多科研基金要求项目申请书包含明确的数据管理计划<sup>[12-14]</sup>。在此背景下,科研人员在数据的共享、发表、重用以及研究数据管理等方面的数据实践也成为研究的热点<sup>[15]</sup>。然而,现有研究较多关注宏观问题,以科研人员个体为中心、从微观的角度考察其在特定任务和目标下的数据行为的研究较少。

宏观政策制度及数据基础设施之外,在第四科研范式中,为科研人员提供获取、组织、分析、可视化、发布和共享数据的工具也很重要<sup>[1,16]</sup>。对数据行为进行深入探究,才能为工具设计带来更实用的启示。虽然已有数据实践研究涉及到对科研人员在特定数据活动中的流程、态度等行为层面的分析<sup>[17-20]</sup>,但学术界尚未明确、系统地提出“数据行为”的概念。缺乏这一概念基础和理论框架,阻碍了对数据行为的有效研究和深入探索。数据行为与信息行为在需求识别、检索、选择、使用等任务和活动上有一定相似之处<sup>[18,21-22]</sup>,信息资源管理学科长期关注数据、文献等知识资源的组织、管理和服务,同时关注用户对知识资源的获取和使

\* 本文系国家自然科学基金面上项目“在线知识社群中信息行为对群体认知形成的影响研究”(项目编号:72174014)和国家自然科学基金国际重点合作项目“中国儒家学术史知识图谱构建研究”(项目编号:72010107003)研究成果之一。

作者简介:李文琦,博士研究生;张鹏翼,长聘副教授,通信作者,E-mail:pengyi@pku.edu.cn。

收稿日期:2022-05-09 修回日期:2022-10-20 本文起止页码:29-40 本文责任编辑:徐健

用行为,因此在连接数据与用户数据行为研究中具有天然的优势。信息行为研究中的概念体系和理论模型经过了几十年实证研究的不断验证和完善,可以作为数据行为概念框架构建的参考。

在此背景下,本文选择数据行为较为密集的科研语境作为出发点,提出两个研究问题:①如何界定“数据行为”及其相关概念?②数据行为的具体类型和影响因素有哪些?

针对以上研究问题,本文综合目前与数据行为相关的研究,借鉴信息行为领域的概念及理论模型,初步构建科研语境中的数据行为模型和概念框架,其目标为:①为理解科研人员数据行为提供一个统一的、个体行为视角的概念框架;②为探究科研人员数据行为中的需求、动机以及具体活动和行为特征提供一个整体的分析框架;③为数据政策、数据管护、数据基础设施建设以及数据工具设计等相关研究提出重点关注方向和研究问题;④为培养科研人员的数据素养、提升数据管理人员的专业技能提供理论基础。

## 2 研究设计

理解“数据”的性质和特征是探究数据相关的行为和活动的前提<sup>[23]</sup>。因此,本文首先梳理数据的多种定义,分析不同学科研究数据的性质差异性,进而阐述

由此产生的数据收集、分析等行为活动的复杂性。随后,本文对现有的数据行为和活动相关的主要概念和研究领域进行分析,在此基础上论述“数据行为”概念的必要性。进而本文借鉴信息行为领域的相关定义和理论模型,探讨提出数据行为概念框架的可行性,界定数据行为相关概念。最后构建科研语境中的数据行为概念框架并提出未来该领域可能的研究方向。

基于上述研究思路,本文主要在 Web of Science、中国知网和 Google Scholar 中检索并筛选了高质量的理论研究以及相关性较大的实证研究,通过引文和关键作者追踪进行文献补充,最终选择了 91 篇文献。其中,为了阐述数据的性质及特征,本文主要检索了在 e-Science 和开放科学等背景下各科研管理和资助机构对数据的定义及对数据性质的阐述,同时收集了不同学科研究方法、科研实践方面的文献以探究数据相关活动的复杂性。为了分析现有数据行为、活动相关研究的不同视角,探究数据行为概念提出的必要性,主要通过数据实践、数据重用、数据共享、科研数据管理、数据管护及数据生命周期等关键词检索了图书馆学情报学相关领域的相关文献。为了借鉴信息行为领域的研究论述框架构建的可行性,主要参考信息行为理论专著,并以信息行为、信息搜寻及信息利用等关键词检索了科研语境中的相关研究。具体文献收集流程及数量如图 1 所示:

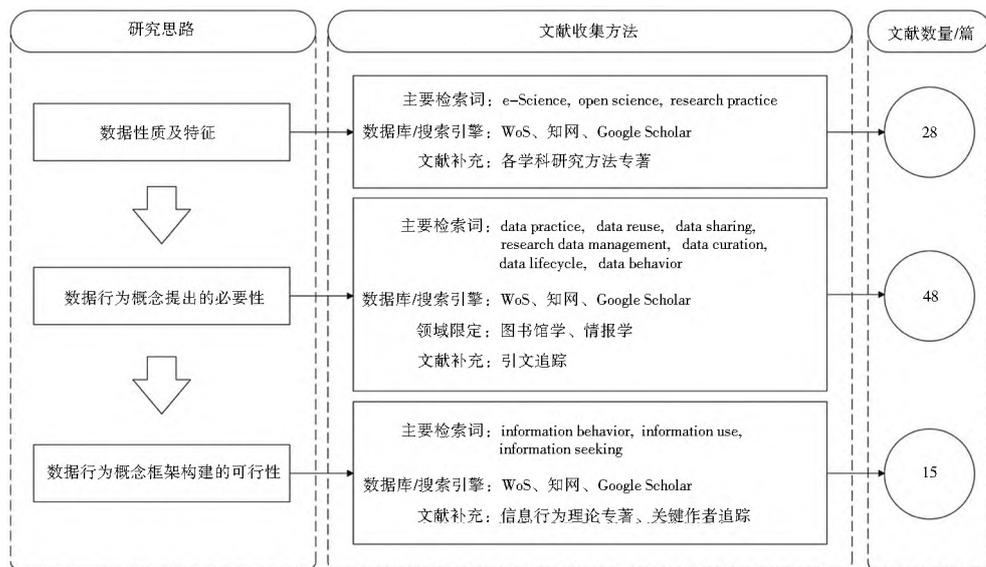


图1 文献收集流程

## 3 科研数据性质及特征

### 3.1 数据定义的多样性

数据在各学科和研究领域中广泛存在,其概念内

涵和外延因各领域本体论、认识论和方法论的不同而呈现差异<sup>[23]</sup>。目前学界对于科研语境中的数据并没有统一、明确的定义<sup>[24]</sup>。国内外诸多科研管理和资助机构分别从数据的用途和覆盖类型等方面对数据进行

了定义,具体如表 1 所示:

表 1 部分科研管理和资助机构对研究数据的定义

来源	定义
国务院办公厅 《科学数据管理 办法》 <sup>[25]</sup>	科学数据主要包括在自然科学、工程技术科学等领域,通过基础研究、应用研究、试验开发等产生的数据,以及通过观测监测、考察调查、检验检测等方式取得并用于科学研究活动的原始数据及其衍生数据
美国国家科学基 金会 <sup>[13]</sup>	科学界普遍接受的,作为验证研究结果所必需的事实材料记录
美国国家人文 基金 <sup>[12]</sup>	在进行研究过程中产生或收集的材料
欧盟委员会 zon 2020 <sup>[14]</sup>	Hori- 研究数据是指为了审查、考虑,并作为推理、讨论或计算的基础所收集到的信息,特别是事实或数字
英国经济和社会 研究委员会 <sup>[26]</sup>	作为研究输入或输出的、与科研人员相关或感兴趣的信 息。是为了在研究项目过程中进行分析而收集或生成的 原始数据,或者是从现有来源获取的研究材料

从这些定义中可以概括出数据的一些性质:

- 是验证研究结果所必需的材料;
- 既包括收集到的用来分析的输入性材料,也包括研究过程中的相关产出;
- 可以是在研究过程中产生或收集的新的原始数据,也可以是从现有来源获取到的研究资料。

C. L. Borgman 认为,观测值、实验记录以及文献等实体是否能作为数据取决于科研人员的学术判断或利益共同体的管理决策,不同的组织机构对数据进行定义的侧重点不同<sup>[27]</sup>。为了分析不同学科对数据的理解以及数据实践上的异同,她将数据定义为出于研究或学术目的而被用作描述或解释现象的证据的实体<sup>[23]</sup>。本文对“数据行为”的论述是基于这一定义展开的。

上述研究中,“数据”和“科研数据”经常交替出现,本文统一用“数据”来指代科研语境中的数据。此外,“数据集”(dataset)也常与“数据”交替使用。一般来说,数据指代数据集的内容,且数据集具有以下特点:①对内容的聚合;②内容的相关性;③为了明确的科研目的而创建<sup>[28]</sup>。虽然数据集的概念在数据共享、重用和引用的研究中经常被提到,但研究数据作为构成数据集的内容,更适合本文对数据和数据行为的阐述和分析。

### 3.2 数据性质的差异性

理解数据的性质是探究围绕数据展开的收集、分析、管理等活动的基础<sup>[23]</sup>。数据的性质包含多种维度,其中最基本的是数据形式和数据来源。部分数据定义列举了不同的数据形式,如欧盟委员会提到研究数据包括统计数据、实验结果、测量结果、实地工作观察结果、调查结果、访谈记录和图像等<sup>[14]</sup>。美国国家

人文基金中给出的人文数据的示例包括引文、软件代码、算法、数字工具、文档、数据库、地理空间坐标(例如来自考古挖掘)、报告和文章等<sup>[12]</sup>。根据国家科学委员会的报告,数据涉及数字、文本、多媒体、模型或者实物等多种形式,可以通过观测、实验或计算 3 种途径产生<sup>[29]</sup>。学科核心问题和科研范式的差异导致其开展研究所依赖的数据性质大相径庭<sup>[30]</sup>。以数据形式为例,人文学科多以实物或者文本为主,而科学数据则多是数字和模型;从数据来源上看,科学研究中大多是观测数据或实验数据,而人文研究几乎不涉及实验数据。

为了更好地概括和区分不同知识领域中的数据特性或科研人员的行为,本研究以“元学科”为分析单元<sup>[31]</sup>,沿用 C. L. Palmer 等<sup>[32]</sup>和 C. L. Borgman<sup>[23]</sup>的三类元学科划分,即科学、社会科学及人文,从数据形式和数据来源两个维度对数据性质进行了梳理,见表 2。

科学研究包括对自然界各类事物和现象的研究,其数据是自然界的实物样本以及对样本的测量、对现象的观测记录或仿真模型<sup>[23]</sup>。科学研究中的数据产生于科研团体内部,也只服务于科研目的<sup>[33]</sup>。随着开放科学的推进,已有的外部数据成为重要的科学研究数据来源,例如公开数据仓储库(如斯隆数字寻天项目)<sup>[34]</sup>、学术发表中公开的数据集、实验方法和计算模型以及科学研究中特有的规范数据(如基因序列和化学结构)等。

社会科学的研究对象是人类行为及社会现象,其数据是对研究对象的行为活动的记录,涉及各种形式和来源的数据<sup>[35]</sup>。既包括科研人员对观察或实验中的人的行为、语言甚至是表情等生理反应的记录和测度,也包括人类社会活动的自然产物或是为了服务于社会组织活动而收集的数据,如政府报告、经济统计以及社交媒体上的帖子等<sup>[36]</sup>。社会科学中的研究范式与数据的性质直接相关,例如实证主义通常依赖于定量数据,解释主义以定性数据为主,后实证主义则是定性定量数据的结合<sup>[35]</sup>。定量数据常见的来源包括调查问卷、实验数据或数字化行为痕迹数据等,定性数据包括访谈、日志或田野观察的记录等。

人文研究的对象是人类文化和人类记录<sup>[23]</sup>。有学者认为人文研究中没有类似科学研究中的数据,只有来源(sources)<sup>[37]</sup>。在传统人文研究中,数据就是研究对象本身,是用以支持论点的证据<sup>[38]</sup>。任何人类文化的记录都可以作为人文研究数据<sup>[34]</sup>,包括档案资料、出版物或者文化产物等<sup>[39]</sup>。人文研究数据从形式

表2 不同元学科的数据性质

类别		科学	社会科学	人文
数据形式	数字	仪器测量数据;计算数值等	统计数据;量表数据等	-
	文本	观察笔记;描述性测量数据	访谈中受访者的回答;日记;定性分析编码	历史资料;文学作品等文献
	多媒体	自然现象或实验结果的影像记录	观察或实验中对人类行为及反应的影像记录	画作、影视、音乐等创作;人类活动影像记录
	模型	仿真模型	仿真模型	-
	实物	环境样本、标本等	日常物品等	文物等
数据来源	生成或收集新数据	观测	现场记录(机器类型、校准、时间、日期和地点,采集方法);照片或录像;样本和标本等	-
		实验	实验设计(变量、实验程序、材料、仪器等);过程记录;实验结果	-
	计算	建模和模拟	建模和模拟	数字化文本(如OCR);标记文本等
	收集外部已有数据	类型	现有观测数据;可以再现数据的实验设计及模型;规范或参考数据(如基因序列、化学结构)	现有观测数据(如交易和搜索日志、社交媒体内容);可再现数据的模型;档案资料;公共记录或报告(如人口普查、公司记录、经济指数)
	渠道	合作者或其他学者;公共数据存储库	合作者或其他学者;公共数据存储库;图书馆、档案馆	图书馆、档案馆(主要);合作者或其他学者;公共数据存储库

上看呈现出很强的异质性<sup>[40]</sup>,包括文本、图像、视频、音频以及实物等,但较少涉及到数字和仿真模型等形式。从数据来源看,人文研究很少从观测中产生数据<sup>[34]</sup>,也不通过实验获取数据<sup>[36]</sup>。人文研究更依赖外部数据源<sup>[34]</sup>,使用的多是回溯性的、由他人或机构创建、收集和整理的数<sup>[33,41]</sup>。随着数字人文的发展,OCR(光学字符识别)、机器自动标注及知识抽取等数字化方法产生的文本资料也成为了人文研究中的重要数据。在数字人文研究中,数据是对人文研究对象的某些方面的抽象,这种抽象是数字化的、选择性构建的和可计算的<sup>[42]</sup>。

### 3.3 数据活动的复杂性

在数据密集型的第四科研范式及开放科学背景下,科研人员围绕数据展开的活动变得更加丰富和复杂,不仅需要应对大规模的异质性数据的收集、处理和分析,还涉及数据共享、数据重用以及研究数据管理等活动。科研人员的行为特征会影响数据基础设施建设、数据管护以及相应的平台和工具设计<sup>[30,32,43]</sup>。只有深入理解这些特征才能为其提供更完备的设施和服务。不同元学科数据收集和分析等活动的特点也受到学科研究范式及文化的影响<sup>[43]</sup>。

科学研究以往通常依赖望远镜、显微镜、传感器等辅助感官或测量仪器来收集数据<sup>[44]</sup>。随着技术的不断进步,很多科学数据是通过传感器网络等设施实现结构化、自动化采集的<sup>[45]</sup>。对于需要人工收集的数据,为保证准确性和可回溯性,通常需要多个独立观察

者共同完成,并利用照片、录像等作为辅助记录手段。科学研究强调客观性和可重复性,数据收集人员需要严格遵循观测或实验的流程,使用统一的测量标准及精确的描述性词语。这种客观性和程式化的特点允许非专业科研人员进行数据观测<sup>[36]</sup>,例如公民科学(citizen science)领域的研究数据常由经过培训的志愿者收集<sup>[46]</sup>。科学数据的分析大多是定量的<sup>[44]</sup>,特别是大规模异质科学数据很难通过纯人工手段处理分析,一般会使用到软件、计算方法和统计<sup>[23]</sup>,甚至还需要特定的软、硬件设备(如基因测序仪等)。

在社会科学研究中,数据可以由科研人员及其协作团队进行收集,也可以由网络活动追踪软件等工具自动收集。其中定性数据的收集需要一定的研究素养并结合情境选择性构建,因此很难由研究团队以外的人完成。社会科学的数据分析也有规范的方法和流程,例如在定量分析中需要首先建立假设并对假设进行验证,然后拒绝或接受假设并作出可能的理论解释;定性分析的过程则包括数据压缩、数据展示、得出并验证结论等<sup>[47]</sup>。此外,近年来网络分析及机器学习等大数据研究手段也在社会科学中广泛应用<sup>[48]</sup>。社会科学的数据分析大多可使用通用分析软件,如用于统计分析的R、SPSS、SAS,用于社会网络分析的Gephi及用于定性内容分析的NVivo等。此外,社会科学的科研人员对数据共享有更多顾虑,特别是定性数据的共享。这些顾虑包括数据重用时是否会对数据产生错误解读,以及是否会侵犯研究对象的隐私等<sup>[36]</sup>。尽管

如此,数据共享与重用也逐渐在社会科学中发展起来,调查结果、访谈材料等也逐步在科研人员的斟酌和权衡下通过学术发表及公开数据存储库进行共享。

在传统人文研究中,数据主要依靠人工收集,收集人员通常就是研究者本人<sup>[49]</sup>。较为成熟的人文学者通常会维护个人常用的核心文献集作为主要的数据来源<sup>[50]</sup>,人文学者自身的学术专业水平对所收集的数据质量和价值有很大影响。在数字人文研究中,学者需要使用更大规模的数据,因此数据收集和整理更加依赖图书馆、档案馆等机构以及众包志愿者等。人文研究通常是对已有资料的诠释,注重叙述性、判断性的推理和论证而非特定的规范性分析方法。对已知事实的掌握是人文学者的基本要求<sup>[51]</sup>,因此在传统人文研究中,对研究数据的整理、融合、内化甚至是记忆十分重要。此类研究依赖于学者对内容的精细研读,十分耗时<sup>[52]</sup>,因此传统人文研究一般不会涉及到很大的数据规模。而数字人文研究借助正则表达式、机器学习等技术使大规模数据的自动化处理成为可能,同时定量分析、网络分析以及可视化等技术的引入也可以帮助人文学者从大规模数据中发现规律<sup>[53]</sup>。

## 4 数据行为概念框架构建

### 4.1 概念提出的必要性

目前已有较多针对特定类型数据活动(如数据共享和重用等)的研究,主要围绕数据实践(data practice)和数据生命周期(data lifecycle)两个方面展开,但尚未有研究对各类数据行为、活动进行全面的梳理和系统性的定义。

C. L. Palmer 等认为数据实践包括了生成、管理和共享数据等工作,是一种重要的学术原语<sup>[32]</sup>,即所有学科所共有的基本学术活动<sup>[54]</sup>。对数据实践最为全面的梳理来自于 T. C. Chao 等构建的数据实践及数据管护术语框架<sup>[30]</sup>,该框架通过对科研中的各类数据实践以及每种数据实践包含的一系列活动的列举明确了数据实践的外延,包括研究设计、数据管理、数据收集和产生、数据处理、数据分析解释及抽象、数据表示、数据共享、数据引用以及数据发表。以数据处理为例,其所涉及的具体活动包括预处理、数据清洗、数据转化等。不少研究从图书馆和数据管护等视角聚焦于特定学科或元学科中的数据实践<sup>[55-57]</sup>。数据共享与重用<sup>[17,19,21,58-59]</sup>以及研究数据管理<sup>[60-62]</sup>等方面的数据实践也随着开放科学的发展成为研究热点,有些研究分别从不同视角对具体的数据实践类型进行了定义和探

索<sup>[63]</sup>。总体而言,关于数据实践的研究大多是从组织机构、政策以及基础设施及服务角度出发,关注学术研究全局或者某一元学科、学科或学术团体中普遍、广泛的实践及经验,而非关注科研人员的具体情境和任务下的个体行为特征。

数据生命周期是在研究数据管理或管护时常用的概念<sup>[8-9]</sup>。研究数据的生命周期包括研究计划、数据收集、数据处理和分析、数据发布和共享、数据保存、数据发现、数据重用等数据相关的活动<sup>[64,65]</sup>;而数据管护的生命周期模型还包括数据筛选和数据存储等由图书馆员、数据管理人员而非科研人员来完成的活动<sup>[8]</sup>。数据生命周期是以数据为核心,以为数据提供积极、持续的维护和管理为目的而进行的数据活动的梳理,目的是使数据得以长期保存并为现在和未来的数据使用者提供使用价值,而较少涉及到科研人员即数据持有者的数据需求及收集和使用中的行为和体验。

综上所述,数据实践概念关注的是科研团体、组织机构的广泛的、整体性的现象与经验,而数据生命周期则是以数据从产生到被长期保存和使用过程中的真实性、可靠性及有用性为核心焦点。这些研究都较为宏观,缺乏对科研人员个体在特定情境和任务下的具体行为的深入的观察和分析(见图2)。纵观信息行为研究的发展,也经过了系统导向到个人导向的转变,从最初关注信息源、整体信息使用方式以及系统如何服务于人群转向了个体的需求、动机、习惯、行为以及所处的情境,而这种以人为中心的研究也为系统的设计带来了更多启示。相应的,从数据、群体视角转向为个体视角的数据行为研究也将为数据平台、工具的设计以及政策的制定提供洞见。因此,本文在科研语境下提出“数据行为”的概念,即科研人员在具体研究任务驱动下所产生的数据需求以及相应的数据收集及生产、数据处理、分析及解释、数据表示及数据管理等行为,同时也包括科研人员在学术交流过程中出于特定动机而进行的数据共享、发表及引用等行为。

### 4.2 框架构建的可行性

为对数据行为的子概念及相关概念进行完整梳理,本文借鉴目前比较成熟的、同样是以个体为观察分析对象的信息行为相关理论。已有学者将查找数据以及使用不熟悉的数据等行为与信息搜寻行为相联系<sup>[18,21-22]</sup>,王晓光等<sup>[21]</sup>进一步将数据重用的过程与信息行为的 ISP 模型相对应,并认为数据重用行为与信息行为在识别需求、检索、选择、获取、使用等任务和活动上相似,但在目标和具体行动上有所不同。

	数据实践	数据生命周期	数据行为
研究视角	群体	数据	个体
研究对象	科研团体或组织机构广泛的、整体性的现象与经验	数据在整个生命周期中的真实性、可靠性及可用性	科研人员在具体情境、任务下的个体行为
潜在启示	数据基础设施、建设、政策制定、数据管理等	数据管理、数据的长期保存与管护	数据平台及工具设计、政策制定、数据管理等

图2 数据行为相关概念对比

在科研语境下的信息搜寻研究中, D. Ellis 信息搜寻模型<sup>[66]</sup>及其后来的拓展模型对各元学科的学者在信息搜寻过程中的具体活动作出了较完整的梳理和补充<sup>[67]</sup>,包括准备和计划、开始、调查、浏览、链接、鉴别、提取、监控、验证、访问、社交等,后续拓展模型中还包

括信息使用、信息管理等行为。本文将这些信息活动与数据实践术语表<sup>[68]</sup>进行对照分析(见表3),识别信息行为和数据实践中活动的对应关系,如浏览、链接、验证、社交等搜寻行为,以及处理、分析等使用行为。

表3 数据实践术语与信息行为活动类型对照

	数据实践术语		信息行为活动
数据收集	重用已有数据	搜索、链接、浏览、评估数据质量、验证来源、同辈交流推荐	浏览、链接/引文追踪、鉴别/筛选、验证、社交
	收集或生成新数据	实地研究、实验室研究、外部生产	-
数据使用	数据处理	预处理、清洗、转换	整理和编辑
	数据分析及阐释	分类、比较、注释、搜寻模式、可视化、数据挖掘、开发软件和脚本	整合、分析
	数据表示	生成元数据、应用标准、使用标识符、映射、展示、地理参考	-
数据管理	组织、记录、注释	信息管理	
数据共享	存储到公共数据存储平台、按请求分发、合作共享	信息共享	
数据发表	发表为期刊辅助材料、发表至数据期刊、同步发表提交至数据存储平台、在期刊发表中描述数据	-	
数据署名及引用	在数据发表中共同署名、在发表中正式引用	-	

可以看出信息行为和数据实践中的活动类型有很大重合。因此,借鉴信息行为领域的概念及理论模型,特别是科研语境下的信息行为模型,来构建数据行为的框架具有一定可行性。在信息行为研究中,信息需求、信息搜寻行为、信息使用都是信息行为的子类,参考信息行为相关概念的定义<sup>[69-71]</sup>,本文对数据行为的几个主要子概念进行界定和阐述见表4。

数据行为和信息行为在活动类型上有一定相似性,但这些活动的复杂度、重要性及具体特征在两个领域中呈现很大差别。例如,在数据使用中,数据处理包括预处理、数据清洗以及数字化、结构化、数据聚合等数据转化过程,数据分析涉及到分类、比较、标注、可视化、数据挖掘等分析方法。这些活动的复杂度远高于信息使用中的整理、编辑和分析活动。此外,也有一些

特有的数据行为活动,如通过实地研究或实验室研究收集新数据、数据表示、数据发表、数据署名及引用等活动。数据行为的概念框架构建在借鉴信息行为的理论和理论模型和现有数据实践术语的基础上,还需要进一步对数据行为中复杂而丰富的具体活动及其影响因素进行粒度统一的、个体视角的梳理和分析。

### 4.3 数据行为概念框架

结合表4对数据实践术语及其与信息行为活动的对应关系分析和 T. D. Wilson 提出的信息行为模型<sup>[71-72]</sup>,本文构建了个体视角的数据行为模型并将其置入科研语境中梳理概念框架,见图3。

个体数据行为模型中,数据需求是数据行为发生的动机<sup>[73]</sup>。除数据的主题内容外,数据需求还可能涉及数据形式、数据来源,如是否为一手数据等。与信息

表 4 数据行为与信息行为相关概念对比

数据行为相关概念	信息行为相关概念
<b>数据行为</b>	<b>信息行为</b>
科研人员与数据相关的行为的总体,包括在研究任务驱动下产生的数据需求以及数据收集及生产、数据处理、分析及解释、数据表示及数据管理等行为,也包括研究人员在学术交流过程中进行的数据共享、发表及引用等行为	与信息来源和渠道相关的人类行为的总体,包括主动和被动信息搜寻以及信息使用,也包括信息偶遇、信息规避以及信息共享等行为
<b>数据需求</b>	<b>信息需求</b>
当科研人员意识到自己的知识或证据不足以支撑研究目标时产生的状态	当个体意识到自己当前的知识状态无法满足其目标时产生的一种假设状态
<b>数据收集</b>	<b>信息搜寻</b>
科研人员为了实现研究目标而进行的获取数据的活动,既包括产生新的研究数据,也包括搜集外部已有数据	当个体感知到一个有问题情境或信息鸿沟时对获取信息所作出的有意识地努力
<b>数据使用</b>	<b>信息使用</b>
数据使用是指科研人员如何处理和利用收集到的数据并实现科研目的,包括预处理、分析、解释及数据表示等活动	指个体如何处理获得的信息,包括收集过程和结果或效果

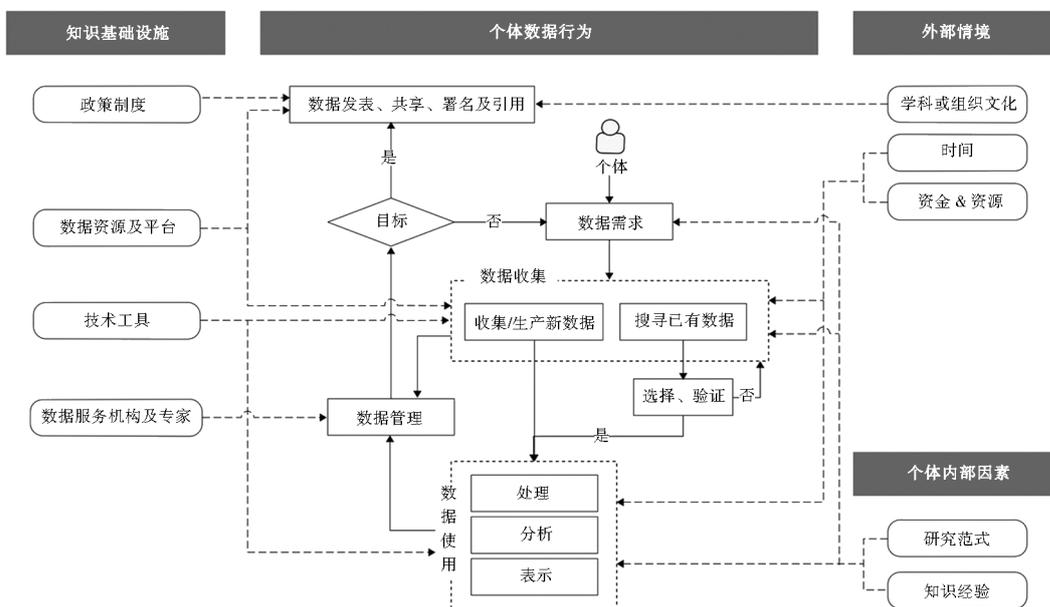


图 3 科研语境中的数据行为框架

需求相比,数据需求更不易表述,科研人员在研究规划与设计过程中对数据需求的认识和表述逐渐清晰,并明确相应的数据收集及分析方法。数据需求背后的动机与其研究的目的密切相关,可以是在探索性研究中形成研究问题,也可以是描述现象、建立或检验理论、预测、支持或反驳观点等。数据需求受到科研人员的数据素养、科研经费及资源投入等因素影响<sup>[74-75]</sup>,如何管理数据需求并有效转化为数据服务也值得图书馆等数据服务机构探索<sup>[76]</sup>。

科研人员在明确数据需求后会产生数据收集行为。数据收集行为既包括收集或生成新的研究数据,也包括搜寻外部已有的数据。其中,新数据的收集则区别于信息获取,涉及到实验、观测或者计算等数据收集或生成方式,以及各类仪器、设备的辅助。已有数据

的收集与信息搜寻相似,包括从外部数据源进行搜寻、选择和获取的过程。数据仓储、元数据检索工具以及期刊出版物等已成为获取开放数据的重要渠道<sup>[77]</sup>。人文学科则依赖大量的现有文献、档案及考古材料,因此数据收集行为也依赖于图书馆、档案馆及博物馆等机构。

科研人员的数据使用行为包括数据处理、分析和表示等。数据使用的方式多样,与学科范式密切相关,同时依赖于技术工具以及科研人员的数据素养<sup>[75,78]</sup>。此外,数据使用还支持科研人员实现科研目的,如描述现象、构建或检验理论、做出预测或者提出研究问题等。在很多人文研究中,数据的作用也体现在科研人员的认知变化,帮助其建立领域事实并启发其思考或批判。

在数据收集和数据处理过程中,科研人员需要进行数据组织、描述及标记等数据管理活动,为后续研究发表及数据共享做准备。专业的数据管理是数据重用的基础,复杂的数据管理通常需与图书馆等专业数据管理服务机构的沟通协作<sup>[79]</sup>。最后,科研人员需要判断经过处理、分析、组织和表示的数据是否满足科研目标,如果没有满足,需要重新明确数据需求并继续数据收集和使用;如果满足,可以对数据进行公开发表、共享,该过程也涉及到数据署名及引用等,所发表和共享的数据则可以进一步为其他科研人员重用。数据共享、重用和引用都很大程度上依赖于研究机构、期刊等制定的政策制度规范<sup>[80]</sup>,同时也会受到不同学科和组织文化的影响<sup>[81]</sup>。

图3中间的个体数据行为模型刻画了数据行为中的活动及其关系。每类数据行为的具体活动、特征以及行为顺序和模式都需要在实证研究中依据个体的目标、具体任务和情境来分析。为结合情境理解数据行为、探讨其影响因素,本文进一步在个体视角的数据行为模型的基础上参考相关研究<sup>[23,82-83]</sup>,引入情境因素及其影响关系,构建科研语境中的数据行为框架。

从宏观视角看,政策制度、数据资源及平台、技术工具、数据服务机构及专家构成了影响和支持各类数据行为的知识基础设施<sup>[23]</sup>。政策制度对数据的发表、共享、署名和引用行为有决定性影响。例如,美国期刊和出版商的数据公开政策准则,使科研人员更直观地了解数据公开的价值,以及发表或引用数据时的具体操作要求<sup>[10]</sup>。数据资源及平台包括数据集、数据存储库及数据存档,决定了科研人员如何、在何处共享以及搜寻已有数据。例如,数据存档对数据的长期存储及迁移的能力,数据存储库在元数据描述、数据溯源以及提升数据的可发现性等方面的数据增值服务,影响科研人员的共享意愿、平台选择及搜寻数据时的体验<sup>[79,84-85]</sup>。技术工具是指数据收集和使用过程中用到的各类软硬件,科研人员在何时使用何种工具是其数据行为的重要特征之一,同时技术工具的可用性及用户体验也可以重塑其数据行为。例如,人文学者为了使用特定的可视化分析工具需要先花费大量时间将其转化为工具可以识别、处理的数据格式<sup>[86]</sup>。数据服务机构及专家包括图书馆、档案馆等机构及数据馆员等专家,其服务会对数据管理产生影响。科研人员对元数据、数据溯源、存档标准等专业数据管理知识的欠缺,会导致数据因记录和表示不够清晰、标准而缺乏长期保存和重用的价值。因此,科研人员能否接触到数

据服务机构和专家并与其高效协作,会影响研究数据管理方面的决策和行为。

影响科研人员数据行为的微观因素包括外部情境和个体内部因素<sup>[83]</sup>。科研语境中的外部情境主要包括学科和组织文化、时间以及资金和资源。其中,学科和组织文化一定程度上影响了科研人员的数据发表、共享意愿以及相应的行为规范,特别是学术团体的协作氛围影响了数据在团体内部和团体之间的共享<sup>[32]</sup>。时间、资金和资源决定了科研项目的规模和研究设计的可行性,因此科研人员的数据需求、数据收集和使用行为都会在一定程度上受到这些因素的限制。个体内部因素主要包括科研人员的研究范式 and 知识经验,其中知识经验不仅包括领域知识和科研经验,还包括数据素养<sup>[77,87]</sup>以及使用技术工具的能力等。这些内部因素对数据需求、数据收集以及数据使用行为都可能产生影响。

## 5 总结与展望

### 5.1 研究贡献与局限

本文通过对研究范式、数据实践以及信息行为等领域的文献梳理,构建科研语境中的数据行为概念框架,主要贡献有以下4点:

首先,提出数据行为的概念,并对数据行为以及数据需求、数据收集行为和数据使用行为3个子概念进行界定,为理解科研人员数据行为提供统一的、个体行为的研究视角。

其次,构建个体视角的数据行为模型,概括了数据行为研究的子领域和一般过程,为探究科研人员数据行为中的具体活动和过程模式的实证研究提供基础的分析框架。

再者,构建科研语境中的数据行为框架揭示了对数据行为产生影响的知識基础设施、科研情境和个体内部因素,可作为后续实证研究提出假设的基础;同时为数据政策制定、数据基础设施建设以及数据工具的设计等领域的研究指出相关数据行为活动。

最后,本文对科研语境中的数据、数据行为及影响情境的梳理和阐述可以为作为科研人员及数据管理人员的数据素养的培训内容参考。

本文主要有两点局限:第一,已有相关研究领域较多、研究视角不统一。本文构建的模型主要基于科研语境,可能不适用日常生活和组织机构中的数据行为,但可作为未来构建通用模型的基础。第二,本文对数据行为的概念界定及其模型构建,主要参照信息行为

领域的理论框架, 仍需实证研究进一步检验和修正。

## 5.2 未来研究展望

在理论方面, 未来研究可以从以下三方面展开:

首先, 对数据行为的各子领域进行深入探究, 理解各类数据行为的特征、驱动该行为的任务和动机等。已有研究更多聚焦于数据共享与重用行为的动机、态度及影响因素等个体数据行为<sup>[19, 81, 88]</sup>, 对科研人员数据需求、数据收集、数据使用及数据管理等行为的研究相对欠缺。未来可以关注这些类型的数据行为, 具体研究问题如: 特定学科的科研人员数据需求是什么; 特定学科的科研人员数据收集或使用行为有哪些具体活动和特征等。

其次, 理解各类数据行为之间的关系、流程和角色分工。本文模型揭示的是一般流程, 但在实际科研过程中, 各类数据行为可能交替、迭代进行并相互影响。未来可以探究在不同学科或科研任务下, 各类数据行为之间的具体过程和相互影响。此外, 科研活动通常在团队中开展, 未来可以探究团队内部和团队间的数据角色分工、协同数据行为和数据共享模式等。

再次, 验证并理解知识基础设施、科研及个体内部情境等因素对数据行为的具体影响。目前在开放科学背景下对数据共享和重用行为的影响因素的研究较多, 尚欠缺对基础的数据收集、使用以及数据管理等行为的影响因素, 特别是个体内部因素的研究。

在实践方面, 数据行为研究可以对不同利益相关者有所启示, 包括科研资助和管理机构、期刊编辑、同行评议专家、图书馆、数据档案、高校、科研机构以及科研人员等。首先, 由科研资助和管理机构、期刊提出的数据管理计划、数据发布要求等政策制度没有充分考虑到各学科科研人员的数据行为的多样性<sup>[23]</sup>, 以及时间、资金等方面的限制。未来对不同学科数据发表、共享、署名及引用行为、数据管理行为及其影响因素的研究可以为科研资助和管理机构、期刊编辑、同行评议专家等数据政策制定和监管者提供参考依据, 也可以为高校及科研机构优化科研评价及激励制度提出建议。其次, 以往学科馆员主要负责文献查新服务及信息素养培训等, 现在还需要提供数据管理、数据分析以及数据可视化等参考咨询服务及数据素养培训<sup>[89-91]</sup>。未来对各学科数据行为的各类数据行为的研究可以为图书馆等机构完善数据服务和科研支持提供启示。

在系统设计方面, 数据行为依赖于各类数据平台及工具的建设, 而数据平台及工具的设计更不能脱离对用户行为的理解。相较于信息搜索工具设计, 目前

尚欠缺对数据平台及工具设计的研究。如何提升数据平台及工具的用户体验, 使其更好地融入科研人员的数据流程, 也是未来数据行为研究需要关注的方向。例如, 对科研人员的数据发表、共享和署名行为、数据管理行为以及数据搜寻行为的研究可以为数据存储与检索平台的设计提供启示, 同时在数据收集、数据处理、分析、可视化以及科研协作等工具的设计上给出建议<sup>[86]</sup>。

## 参考文献:

- [1] GRAY J, LIU D T, NIETO-SANTISTEBAN M, et al. Scientific data management in the coming decade[J]. ACM sigmod record, 2005, 34(4): 34-41.
- [2] SCHROEDER R. Big data: towards a more scientific social science and humanities? [M]. Oxford: Oxford University Press, 2014.
- [3] HEY T, TREFETHEN A. E-Science and its implications[J]. Philosophical transactions: mathematical, physical and engineering sciences, 2003, 361(1809): 1809-1825.
- [4] ATKINS D. Revolutionizing science and engineering through cyberinfrastructure; report of the national science foundation blue-ribbon advisory panel on cyberinfrastructure[R/OL]. [2022-09-03]. <https://www.nsf.gov/cise/sci/reports/atkins.pdf>.
- [5] CRANE G, BABEU A, BAMMAN D. ESscience and the humanities[J]. International journal on digital libraries, 2007, 7(1/2): 117-122.
- [6] UNSWORTH J. Our cultural commonwealth: the report of the American Council of learned societies commission on cyberinfrastructure for the humanities and social sciences[R]. New York: ACLS, 2006.
- [7] COX A M, TAM W W T. A critical analysis of lifecycle models of the research process and research data management[J]. Aslib journal of information management, 2018, 70(2): 142-157.
- [8] HIGGINS S. The DCC curation lifecycle model[J]. International journal of digital curation, 2008, 3(1): 134-140.
- [9] CHRISTOPHERSON L, MANDAL A, SCOTT E, et al. Toward a data lifecycle model for NSF large facilities[C]//Practice and experience in advanced research computing. New York: Association for Computing Machinery, 2020: 168-175.
- [10] NOSEK B A, ALTER G, BANKS G C, et al. Promoting an open research culture[J]. Science, 2015, 348(6242): 1422-1425.
- [11] 陈传夫, 李秋实. 数据开放获取使科学惠及更广——中国开放科学与科学数据开放获取的进展与前瞻[J]. 信息资源管理学报, 2020, 10(1): 4-13.
- [12] NATIONAL ENDOWMENT FOR THE HUMANITIES. Data management plans for NEH office of digital humanities proposals and awards[EB/OL]. [2022-09-20]. <https://www.neh.gov/sites/default/files/inline-files/Data%20Management%20Plans%20202019.pdf>.
- [13] NATIONAL SCIENCE FOUNDATION. NSF's public access plan:

- today's data, tomorrow's discoveries[EB/OL]. [2022-09-20]. <https://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf>.
- [14] EUROPEAN COMMISSION. Open access - H2020 online manual [EB/OL]. [2022-09-20]. [https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm).
- [15] 王知津, 陈芊颖, 韩峰, 等. 我国开放数据研究进展与趋势(1996-2019年)[J]. 信息资源管理学报, 2020, 10(6): 47-59.
- [16] HEY T, TREFETHEN A. The data deluge: an e-Science perspective[M]//BERMAN F, FOX G, HEY T. Grid computing: making the global infrastructure a reality. Chichester: John Wiley & Sons, 2003: 809-824.
- [17] WHITMORE D A. Seeking context: archaeological practices surrounding the reuse of spatial information[D]. Los Angeles: University of California, 2016.
- [18] ROLLAND B, LEE C P. Beyond trust and reliability: reusing data in collaborative cancer epidemiology research[C]//Proceedings of the 2013 conference on computer supported cooperative work. San Antonio: ACM Press, 2013: 435.
- [19] 张潇月, 宋秀芳, 顾立平, 等. 我国科研人员科研数据重用行为影响因素实证研究——以生物学领域为例[J]. 情报学报, 2021, 40(8): 887-902.
- [20] 严炜炜, 张敏. 科研协同中的数据共享与利用行为模式分析[J]. 情报理论与实践, 2018, 41(1): 55-60.
- [21] WANG X, DUAN Q, LIANG M. Understanding the process of data reuse: an extensive review[J]. Journal of the Association for Information Science and Technology, 2021, 72(9): 1161-1182.
- [22] ZIMMERMAN A. Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse[J]. International journal on digital libraries, 2007, 7(1): 5-16.
- [23] BORGMAN C L. Big data, little data, no data: scholarship in the networked world[M]. Cambridge: MIT Press, 2015.
- [24] POOLE A H. The conceptual landscape of digital curation[J]. Journal of documentation, 2016, 72(5): 961-986.
- [25] 国务院办公厅. 国务院办公厅关于印发科学数据管理办法的通知[EB/OL]. [2022-09-20]. [http://www.gov.cn/zhengce/content/2018-04/02/content\\_5279272.htm](http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm).
- [26] ECONOMIC AND SOCIAL RESEARCH COUNCIL (ESRC). ESRC research data policy[EB/OL]. [2022-09-20]. <https://www.ukri.org/wp-content/uploads/2021/07/ESRC-200721-ResearchDataPolicy.pdf>.
- [27] BORGMAN C L, WALLIS J C, MAYERNIK M S. Who's got the data? interdependencies in science and technology collaborations[J]. Computer supported cooperative work, 2012, 21(6): 485-523.
- [28] RENEAR A H, SACCHI S, WICKETT K M. Definitions of dataset in the scientific and technical literature [C]//Proceedings of the American Society for Information Science and Technology. Pittsburgh: John Wiley & Sons, 2010: 1-4.
- [29] SIMBERLOFF D, BARISH B C, DROEGEMEIER K K, et al. Long-lived digital data collections: enabling research and education in the 21st century[R]. Arlington: National Science Board, 2005.
- [30] CHAO T C, CRAGIN M H, PALMER C L. Data practices and curation vocabulary: an empirically derived framework of scientific data practices and curatorial processes[J]. Journal of the association for information science and technology, 2015, 66(3): 616-633.
- [31] FRY J, SPEZI V, PROBETS S, et al. Towards an understanding of the relationship between disciplinary research cultures and open access repository behaviors[J]. Journal of the Association for Information Science and Technology, 2016, 67(11): 2710-2724.
- [32] PALMER C L, TEFFEAU L C, PIRMAN C M. Scholarly information practices in the online environment: themes from the literature and implications for library service development[M]. Dublin: OCLC Research, 2009.
- [33] TRACE C B, KARADKAR U P. Information management in the humanities: scholarly processes, tools, and the construction of personal collections[J]. Journal of the Association for Information Science and Technology, 2017, 68(2): 491-507.
- [34] BORGMAN C L. Data, disciplines, and scholarly publishing[J]. Learned publishing, 2008, 21(1): 29-38.
- [35] PICKARD A J. Research methods in information[M]. 2nd ed. London: Facet Publishing, 2013.
- [36] BORGMAN C L. The conundrum of sharing research data[J]. Journal of the American Society for Information Science and Technology, 2012, 63(6): 1059-1078.
- [37] KILBRIDE W. Saving the bits: digital humanities forever? [M]. Chichester: John Wiley & Sons, 2015: 408-419.
- [38] OWENS T. Defining data for humanities: text, artifact, information or evidence? journal of digital humanities[J/OL]. Journal of digital humanities, 2011, 1(1) [2022-09-20]. <http://journalofdigitalhumanities.org/1-1/defining-data-for-humanists-by-trevor-owens/>.
- [39] BORGMAN C L. The digital future is now: a call to action for the humanities[J/OL]. Digital humanities quarterly, 2009, 3(4) [2022-09-20]. <http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html%20/000077.html>.
- [40] FLANDERS J, MUÑOZ T. An introduction to humanities data curation[EB/OL] [2022-09-20]. <https://guide.dhcurator.org/contents/intro/>.
- [41] 夏翠娟. 面向人文研究的“数据基础设施”建设——试论图书馆学对数字人文的方法论贡献[J]. 中国图书馆学报, 2020, 46(3): 24-37.
- [42] SCHÖCH C. Big? Smart? Clean? Messy? Data in the humanities[J]. Journal of digital humanities, 2013, 2(3): 2-13.
- [43] PALMER C L, CRAGIN M H. Scholarship and disciplinary practices[J]. Annual review of information science and technology,

- 2008, 42(1): 163-212.
- [44] WILSON E B. An introduction to scientific research[M]. New York: Dover Publications, 1990.
- [45] ABDUL GHANI A N. Experimental research methods for students in built environment and engineering[J]. MATEC web of conferences, 2014, 10: 01001.
- [46] SILVERTOWN J. A new dawn for citizen science[J]. Trends in ecology & evolution, 2009, 24(9): 467-471.
- [47] MILES M B, HUBERMAN A M, SALDANA J. Qualitative data analysis: a methods sourcebook[M]. 3rd ed. Thousand Oaks: SAGE, 2014.
- [48] ZHANG J, WANG W, XIA F, et al. Data-driven computational social science: a survey[J]. Big data research, 2020, 21: 100145.
- [49] STONE S. Humanities scholars: information needs and uses[J]. Journal of documentation, 1982, 38(4): 292-313.
- [50] BARRETT A. The information-seeking habits of graduate student researchers in the humanities1[J]. The journal of academic librarianship, 2005, 31(4): 324-331.
- [51] DALLAS C. Humanistic research, information resources and electronic communication[M]// MEADOWS J, BOECKER H. Electronic communication and research in Europe. Luxembourg: European Commission, 1999: 209-239.
- [52] VAN WIERST P, HOFSTEDE S, OORTWIJN Y, et al. Bolvis: visualization for text-based research in philosophy[C]//3rd Workshop on visualization for the digital humanities. Berlin: IEEE, 2018.
- [53] MORETTI F. Graphs, maps, trees: abstract models for a literary history[M]. London: Verso, 2005.
- [54] UNSWORTH J. Scholarly primitives: what methods do humanities researchers have in common, and how might our tools reflect this[C]//Symposium on humanities computing: formal methods, experimental practice. London: King's College, 2000.
- [55] WELLER T, MONROE-GULICK A. Understanding methodological and disciplinary differences in the data practices of academic researchers[J]. Library hi tech, 2014, 32(3): 467-482.
- [56] THOEGERSEN J L. "Yeah, I guess that's data": data practices and conceptions among humanities faculty[J]. Portal-libraries and the academy, 2018, 18(3): 491-504.
- [57] GREGORY K, GROTH P, COUSIJN H, et al. Searching data: a review of observational data retrieval practices in selected disciplines[J]. Journal of the Association for Information Science and Technology, 2019, 70(5): 419-432.
- [58] ZHU Y. Open-access policy and data-sharing practice in UK academia[J]. Journal of information science, 2020, 46(1): 41-52.
- [59] 李晓, 曲建升, 靳军宝. 科研人员数据重用意愿、行为及满意度影响因素研究: 一项元分析[J]. 图书馆杂志, 2022, 41(7): 128-138, 113.
- [60] REICHMANN S, KLEBEL T, HASANI-MAVRIQI I, et al. Between administration and research: understanding data management practices in an institutional context[J]. Journal of the Association for Information Science and Technology, 2021, 72(11): 1415-1431.
- [61] HUANG Y, COX A M, SBAFFI L. Research data management policy and practice in Chinese university libraries[J]. Journal of the Association for Information Science and Technology, 2021, 72(4): 493-506.
- [62] 唐燕花. 高校科研数据管理服务实践研究及建议[J]. 图书情报工作, 2016, 60(24): 130-138.
- [63] 张潇月, 顾立平, 胡良霖. 国内外开放科研数据重用困境解决措施述评[J]. 图书馆, 2021(3): 80-89.
- [64] UK DATA SERVICE. Research data management[EB/OL]. [2022-09-20]. <https://ukdataservice.ac.uk/learning-hub/research-data-management/>.
- [65] BUDDENBOHM S, CRETIN N, DIJK E, et al. State of the art report on open access publishing of research data in the humanities[R/OL]. [2022-09-03]. <https://shs.hal.science/halshs-01357208>.
- [66] ELLIS D. A behavioural model for information retrieval system design[J]. Journal of information science, 1989, 15(4/5): 237-247.
- [67] SAVOLAINEN R. Contributions to conceptual growth: the elaboration of Ellis's model for information-seeking behavior[J]. Journal of the Association for Information science and Technology, 2017, 68(3): 594-608.
- [68] CHAO T C, CRAGIN M H, PALMER C L. Data practices and curation vocabulary[EB/OL]. [2022-09-20]. <http://hdl.handle.net/2142/44032>.
- [69] PETTIGREW K E, FIDEL R, BRUCE H. Conceptual frameworks in information behavior[J]. Annual review of information science and technology, 2001, 35: 43-78.
- [70] CASE D O, GIVEN L M. Looking for information: a survey of research on information seeking, needs, and behavior[M]. 4th ed. Bingley: Emerald, 2016.
- [71] WILSON T D. Human information behavior[J]. Informing science, 2000, 3(2): 49-56.
- [72] WILSON T D. Models in information behaviour research[J]. Journal of documentation, 1999, 55(3): 249-270.
- [73] 胡媛, 艾文华, 胡子祎, 等. 高校科研人员数据需求管理影响因素框架研究[J]. 中国图书馆学报, 2019, 45(4): 104-121.
- [74] 沈玖玖, 王志远, 戴家武, 等. 基于扎根理论的科研数据需求及影响因素分析[J]. 情报杂志, 2019, 38(4): 175-180, 160.
- [75] 胡卉, 吴鸣. 嵌入科研工作流与数据生命周期的数据素养能力研究[J]. 图书与情报, 2016(4): 125-137.
- [76] 陈媛媛, 柯平. 高校图书馆科研数据服务研究综述[J]. 图书

- 馆工作与研究, 2017(10): 17-23, 30.
- [77] 黄如花. 面向高质量发展的数据素养教育[J]. 图书馆建设, 2020(6): 26-29.
- [78] 孟祥保, 常娥, 叶兰. 数据素养研究: 源起、现状与展望[J]. 中国图书馆学报, 2016, 42(2): 109-126.
- [79] 李晓辉. 图书馆科研数据管理与服务模式探讨[J]. 中国图书馆学报, 2011, 37(5): 46-52.
- [80] 邱均平, 何文静. 科学数据共享与引用行为的相互作用关系研究[J]. 情报理论与实践, 2015, 38(10): 1-5.
- [81] 毕达天, 曹冉, 杜小民. 人文社科科学数据共享意愿影响因素研究——基于同辈压力视角[J]. 情报资料工作, 2020, 41(4): 67-76.
- [82] COURTRIGHT C. Context in information behavior research[J]. Annual review of information science and technology, 2007, 41(1): 273-306.
- [83] FOSTER A. A nonlinear model of information-seeking behavior[J]. Journal of the American society for information science and technology, 2004, 55(3): 228-237.
- [84] 王芳, 赵洪, 马嘉悦, 等. 数据科学视角下数据溯源研究与实践进展[J]. 中国图书馆学报, 2019, 45(5): 79-100.
- [85] 张静蓓, 吕俊生, 田野. 国外数据共享行为影响因素研究综述[J]. 图书情报工作, 2014, 58(4): 136-142.
- [86] GIVEN L M, WILLSON R. Information technology and the humanities scholar: documenting digital research practices[J]. Journal of the Association for Information Science and Technology, 2018, 69(6): 807-819.
- [87] 秦小燕, 初景利. 面向我国科研人员的科学数据素养能力评价研究[J]. 情报理论与实践, 2020, 43(2): 21-27.
- [88] JOO S, KIM S, KIM Y. An exploratory study of health scientists' data reuse behaviors: examining attitudinal, social, and resource factors[J]. Aslib journal of information management, 2017, 69(4): 389-407.
- [89] 顾立平, 张满月. 开放科学环境下数据馆员的实践探析[J]. 图书情报知识, 2020(2): 60-74, 112.
- [90] 闫雪. 国外数据馆员的岗位职责与任职能力研究[J]. 情报科学, 2021, 39(1): 163-168.
- [91] 顾立平. 数据治理——图书馆事业的发展机遇[J]. 中国图书馆学报, 2016, 42(5): 40-56.

## 作者贡献说明:

李文琦: 负责文献调研、模型构建、初稿撰写;

张鹏翼: 负责研究设计、模型构建、论文修改。

## The Conceptualization and Model Construction of Data Behaviors

Li Wenqi Zhang Pengyi

Department of Information Management, Peking University, Beijing 100871

**Abstract:** [Purpose/Significance] With the data-intensive paradigm shift, data has become the foundation of various scholarly activities. Meanwhile, data management and data sharing is gaining more attention from policy makers, research institutions, data service providers as well as researchers. Existing studies have covered topics on data related activities, researchers' attitudes, and behaviors in research context. Yet the concept of data behavior is not explicitly defined and framed. [Method/Process] This paper drew on literatures to analyze the characteristics of research data, illustrate the necessity to propose the concept of data behavior, proposed the concept of data behavior with an individual perspective and develop the data behavior model and concept framework by referring to the theories of information behavior field. [Result/Conclusion] The data behavior model is developed to describe common patterns and process of data behaviors including data needs, data collection, data use, data management and data publishing, sharing, attribution and citation with an individual perspective. A conceptual framework of data behavior in research context encompassing various contextual elements is proposed to reveal potential factors that may influence researchers' data behaviors, which include knowledge infrastructure, research contexts and researchers' internal contexts. The model can provide theoretical basis and practical implications for empirical research of data behavior, data policy formulation, data infrastructure construction and data tool design.

**Keywords:** data behavior research data data practice data-intensive research information behavior